

RANDOM SAMPLING AS A BUILT-IN FUNCTION FOR DATABASE

ADMINISTRATION AND REPLICATION

Abstract

A database management system and method for administration and replication

5 having a built-in random sampling facility for approximation partition analysis on very large
databases. The method utilizes a random sampling algorithm that provides results accurate to
within a few percentage points for large homogeneous databases. The accuracy is not affected
by the size of the database and is determined primarily by the size of the sample. The system and
method for approximate partition analysis reduces the time required for an analysis to a fraction
10 of the time required for an exact analysis. The database management system is configured with
the random sampling facility built-in thereby enabling even greater efficiency by reducing
communication overhead between an analysis program and the database management system to
a fraction of the overhead required when sampling is performed by a separate analysis program.
The reduction in time thereby permits frequent and timely analyses for replication and
15 administration of database partitions.

WFAY_SHARPE_11VOL2IDATA\BJWDATA\2001VMJE\IBM02APP.doc